Exercise 1. Read in the dataset “Pollution.xls”, which we used in a previous lecture.

1) Create a new variable that takes on the value “big” if the population of the city is greater than 500k and “not big” otherwise.
2) Provide summaries of temperature and SO2 by the levels of the new variable. Do temperature and SO2 seem to depend on the fact that the city is big / small?
3) Plot the relationship between population and SO2 and the relationship between temperature and SO2. Explain what you see.

Exercise 2. The dataset “midterms.csv” contains approval ratings and changes in seats in the US Congress after 18 midterm elections.

1) Create a plot to summarize the relationship between approval ratings, change in seats, and party affiliation. Add a line at “approval rating = 50”. Explain what you see.
2) The change in seats in the 2018 midterm election is -38 for the GOP (-40 House, +2 Senate seats). Make a plot that tracks changes in seats over time. Can you notice any patterns? [Hint: Use “SERIES”; you can find examples of usage in the document “Using PROC SGPLOT for quick, high-quality graphs” on the course website.] Add a line at “change in seats = -38” with the label “2018 midterms.” Is the result “surprising”, given the historical data?
3) Does the distribution of approval ratings seem to depend on party affiliation? Explain why or why not.
4) Create a variable that takes on the values “passed” if the approval rating is greater than 50 and “failed” otherwise.
    a. Plot the distribution of changes in seats by the levels of the new variable.
    b. Using PROC TABULATE, create a table that shows changes in seats by the levels of the new variable and party affiliation.
    c. Using parts a. and b. and more figures / tables as necessary, describe the relationship between party affiliation, change in seats, and the new variable.

Exercise 3. The files “before.csv” and “after.csv” have math scores before and after a math bootcamp (on a scale from 0 to 115).

1) Merge the scores in “before.csv” and “after.csv” by student ID (“id”). If there are any unmatched observations, identify them, briefly explain their characteristics, and drop them.
2) Analyze the data. Would you recommend going to the math bootcamp?